10-16-00   A

Attorney Docket No. Semio-991131

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

* * *

UTILITY APPLICATION TRANSMITTAL LETTER

ASSISTANT COMMISSIONER FOR PATENTS
Box Patent Application
Washington, D.C. 20231

Sir:

Transmitted herewith for filing is the patent application of:

INVENTOR(s):   Claude VOGEL

FOR: SYSTEM AND METHOD FOR PARSING A DOCUMENT

This Application Claims Priority From U.S. Application Serial No. 09/288,994, Filed April 9, 1999.

Enclosed are:

[14] Pages of specification
[3] Pages of claims
[1] Page of abstract
[17] Sheets of informal drawings
[x] Unexecuted Declaration and Power of Attorney
[x] Return Postcard

All future correspondence should be addressed to:

Timothy W. Lohse
GRAY CARY WARE & FREIDENRICH LLP
3340 Hillview Avenue
Palo Alto, CA 94304
(650) 320-7400

Respectfully submitted,

GRAY CARY WARE & FREIDENRICH LLP

Dated: 10/10/2000
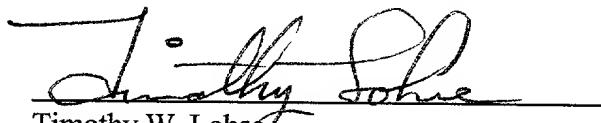
Timothy W. Lohse
Reg. No. 35,255
Attorney for Applicant

GRAY CARY WARE & FREIDENRICH LLP
3340 Hillview Avenue
Palo Alto, CA 94304
Telephone: (650) 320-7400

SYSTEM AND METHOD FOR PARSING A DOCUMENT

Background of the Invention

This invention relates generally to a system and method for processing a document and in particular to a system and method for identifying a plurality of phrases within the document

5   which indicate the context of the document.

Various factors have contributed to the extensive storage and retrieval of textual data information using computer databases. A dramatic increase in the storage capacity of hard drives coupled with a decrease in the cost of computer hard drives, and increases in the transmission speed of computer communications have been factors. In addition, the increased processing

10   speed of computers and the expansion of computer communications networks, such as a bulletin board or the Internet, have been factors. People therefore have access to the large amounts of textual data stored in these databases. However, although the technology facilitates the storage of and the access to the large amounts of textual data, there are new problems that have been created by the large amount of textual data that is now available.

15   In particular, a person trying to access textual data in a computer database having a large amount of data needs a system for analyzing the data in order to retrieve the desired information quickly and efficiently without retrieving extraneous information. In addition, the user of the system needs an efficient system for condensing each large document into a plurality of phrases (one or more words) which characterize the document so that the user of the system can

20   understand the document without actually viewing the entire document. A system for

condensing each document into a plurality of key phrases is known as a parsing system or a parser.

In one typical parser, the parser attempts to identify phrases which are repeated often within the document and identifies those phrases as being key phrases which characterize the

5   document. The problem with such a system is that it is very slow since it must count the repetitions of each phrase in the document. It also requires a large amount of memory. As the amount of data to be parsed increases, the slow speed of this parser becomes unacceptable. Another typical parser performs a three step process to identify the key phrases. First, each word in the document is assigned a tag based on the part of speech of the word (i.e., noun, adjective,

10   adverb, verb, etc.) and certain parts of speech, such as an article or an adjective, may be removed from the list of phrases which characterizes the document. Next, one or more sequences of words (templates) may be used to identify and remove phrases which do not add any understanding to the document. Finally, any phrase which is an appropriate part of speech and does not fall within one of the templates is accepted as a key phrase which characterizes the

15   document. This conventional parser, however, is also slow which is unacceptable as the amount of data to be parsed increases.

In all of these conventional parser systems, the parser attempts to break the document down into smaller pieces based on the characteristics (frequency of repetition or part of speech) of the particular words in the document. The problem is that language generally is not that easily

20   classified and therefore the conventional parser does not accurately parse the document or requires a large amount of time to parse the document. In addition, the conventional parser

systems are very slow because they all attempt to use complex characteristics of the language as

a method for parsing the key phrases out of the document. These problems with conventional

parsers becomes more severe as the number of documents which must be parsed increases.

Today, the number of documents which must be parsed is steadily increasing at a tremendous

5   rate due to, among other things, the Internet and the World Wide Web. Therefore, these

conventional parsers are not acceptable. Thus, it is desirable to provide a parsing system and

method which solves the above problems and limitations with conventional parsing systems and

it is to this end that the present invention is directed.

Summary of the Invention

10   A parser system and method in accordance with the invention is provided in which the

break characters within a sentence or a paragraph are used to parse the document into a plurality

of key phrases. The parser system in accordance with the invention is very fast and does not

sacrifice much accuracy for the speed. The break characters within the document may include

punctuation marks, certain stop words and certain types of words such as verbs and articles. The

15   parser system may include a buffer which receives one or more words before it receives a break

character. When the buffer receives a break character, the parser may determine whether the

phrase before the break character is saved based on the type of break character. In particular, if

the break character is a punctuation mark, the parser may keep the one or more words before the

break character as a key phrase. If the break character is another type of character, the phrase

20   before the break character may or may not be saved. Once the fate of the phrase has been

determined, the buffer is flushed and the next sequence of one or more words is read into the

buffer so that it may also be parsed. In this manner, a plurality of phrases in the document may be rapidly extracted from the document based on the break characters within the sentences and paragraphs of the document.

5   The parser system in accordance with the invention may also be used to parse various different foreign languages into phrases provided that the rules database includes rules that are applicable to the particular foreign language. In particular, each foreign language may have slightly different syntax or characters (in the case of Asian languages or Arabic, for example) so that the rules must reflect those syntactic and character differences.

10   Thus, in accordance with the invention, a system for parsing a piece of text into one or more phrases which characterize the document is provided. The system comprises a buffer for reading one or more words from the piece of text into the buffer and a parser for identifying a phrase contained in the buffer, the phrase being a sequence of two or more words in between break characters. The parser further comprises means for determining the type of break character that follows the identified phrase and means for saving a key phrase from the buffer based on the

15   determined type of break character. The key phrases are stored in a database.

          In accordance with another aspect of the invention, the parsing method may include a two-pass process wherein phrases are extracted from the piece of text as described above. During the second pass, all of the occurrences of the extracted phrases in the piece of text are retrieved. The second pass ensures that phrases that were not extracted at each location in the

20   piece of text may still be retrieved.

Brief Description of the Drawings

Figure 1 is a block diagram of a text processing system;

Figure 2 is a block diagram of a parsing system in accordance with the invention;

Figure 3A is a flowchart illustrating a two-pass parsing method in accordance with the invention;

Figure 3B is a flowchart illustrating more details of the extracting phrases step of the parsing method shown in Figure 3A;

Figure 4 is an example of a document to be parsed by the parsing system in accordance with the invention;

Figures 5A - 5L are diagrams illustrating the operation of the parsing buffer in accordance with the invention on the document shown in Figure 4;

Figure 6 is a diagram illustrating a piece of Japanese text; and

Figure 7 is a diagram illustrating the Japanese phrases extracted from the Japanese text of Figure 6 in accordance with the invention.

Detailed Description of a Preferred Embodiment

The invention is particularly applicable to a system for parsing English language documents and it is in this context that the invention will be described. It will be appreciated, however, that the system and method in accordance with the invention has greater utility, such as

to other languages and to various different pieces of textual data. To better understand the

invention, a text processing system will now be described.

Figure 1 is a block diagram of a text processing system 10. The text processing system

10 may include a parser system 12, a clusterizer 14, a map generator 16 and a database (DB) 18.

5   The text processing system may receive one or more pieces of text , such as stories, press

releases or documents, and may generate a map graphically showing the relationships between

the key phrases in the document. Each piece of text may be received by the parser system 12

which processes each piece of incoming text and generates one or more key phrases for each

piece of text which characterizes the piece of text. The key phrases may be stored in the database

10   18. The details about the parser system will be described below with reference to Figures 2- 5.

Once the key phrases are extracted from each piece of text, the clusterizer 14 may generate one

or more clusters of the key phrases based on the relationships between the phrases. The clusters

generated may also be stored in the database 18. The map generator 16 may use the generated

clusters for the pieces of text in the database in order to generate a graphical map showing the

15   relationships of the key phrases within the various pieces of text in the database to each other so

that a user of the system may easily search through the database by viewing the key phrases of

the pieces of text. More details about the clusterizer and map generator are disclosed in co-

pending U.S. patent application serial no. 08/801,970 which is owned by the assignee of the

present invention and is incorporated herein by reference. The text processing system may be

20   implemented in a variety of manners including a client/server type computer system in which the

client computers access the server via a public computer network, such as the Internet. The

parser, the clusterizer and the map generator may be software applications being executed by a central processing unit (not shown) of the text processing system 10. Now, the parser system 12 in accordance with the invention will be described in more detail.

Figure 2 is a block diagram of the parsing system 12 in accordance with the invention.

5   The parsing system 12 may include a buffer 20, a parser 22 and a rules database (rules DB) 24. The buffer may store one or more words of the incoming piece of text, which may be a document, which are analyzed by the parser 22 using the rules contained in the rules DB 24. The output of the parser system 12 is one or more phrases (each phrase containing one or more words) which characterize the document being parsed. In particular, the parser may separate

10   phrases in the document based on break characters within the document in accordance with the invention. In more detail, one or more words may be read into the buffer from the document until a break character is identified. Thus, the parser system 12 identifies phrases which are between break characters. Then, based on the type of break character, the phrase may be saved as a key phrase or deleted. The parser system 12, for example, may be implemented as one or

15   more pieces of software being executed by a microprocessor (not shown) of a server computer which may be accessed by a plurality of client computers over a computer network, such as the Internet, a local area network or a wide area network. The parser 22 advantageously rapidly extracts key phrases from a piece of text using break characters. The break characters in accordance with the invention will now be described.

20   The break characters may include an explicit break, such as a punctuation mark, numbers, words containing numbers, and stop words. The stop words may be further classified as soft stop

words or a hard stop words. Each of these different break characters will now be described. The explicit break characters may include various punctuation symbols, such as a period, a comma, a semicolon, a colon, an exclamation point, right or left parenthesis, left or right square brackets, left or right curly braces, a return character or a line feed character. The stop characters may be a

5   generated list or it may include a slash (/) and an ampersand symbol (@). A separator may be defined as digits, letters, foreign characters, break characters, apostrophes, dashes and other stop characters. The various words in a piece of text may be categorized as articles, connectors, hard and soft stop characters, linguistic indicators, a syntactic categories such as nouns, verbs, irregular verbs, adjectives and adverbs.

10   In parsing the characters in the piece of text, separators may always be added to a phrase. A apostrophe or dash at the beginning of a word is treated as a break character (see below), an apostrophe or dash at the end of a word is also treated as a break character and a word with an apostrophe or dash in the middle of the word is added to the phrase in the buffer. All stop characters and breaks are treated as stop characters and breaks as described below. At the word

15   level of parsing, proper nouns are retained by testing for an upper case letter at the first character of the word. In addition, all words with only upper case letters and numeric words are kept in the buffer. Optionally, a numeric string may be classified and treated as a stop character. The following are mandatory word level parsing rules. First, the word following as possessive "s" may be deleted. For example, as the sentence "The cat's paw is wet." is parsed in accordance

20   with the invention, "the" is deleted and "cat" is put into the buffer and then deleted when the break character (the aprostrophe) is detected. The apostrophe is deleted because it is punctuation

and then the next character to parse is the possessive "s" after the apostrophe which is deleted

along with the word "paw" since it follows the possessive "s". Connector words appearing at the

beginning of a phrase are also deleted although a connector word followed by "The" is kept in

the buffer. For a hard stop character, the last phrase connected to the hard stop character is

5   deleted and the remaining buffer is processed. A soft stop character may be treated as a break

character. A repeated character is treated as a stop character.

To further remove unwanted words for parsing. some optional phrase level parsing rules

may be used. In particular, phrases longer than a predetermined length, such as six words, may

be deleted, a phrase with all upper case words may be deleted and a phrase with all numeric

10   words may be deleted. All of the above parsing rules may be stored in the parsing rules database

24 shown in Figure 2. Now, the details of the parser system 12 will now be described with

reference to Figures 3A and 3B.

Figure 3A is a flowchart illustrating a two pass parsing process 30 in accordance with the

invention. In particular, during a first pass 40, one or more phrases are extracted from a piece of

15   text using the hard and soft stop words as described below with reference to Figure 3B. The first

pass thus extracts noun phrases. For example, if a piece of text includes, "The big frog and the

kangaroo jumps down.", the first pass extracts the phrase "big frog", but not "kangaroo jumps"

as described below. During a second pass 41, all extracted phrases are retrieved from the piece

of text. In particular, the occurrence of each extracted phrase in the piece of text may be

20   retrieved from the piece of text. For example, assume that a piece of text contains the fragments,

"The software bugs on…" and "software bugs are…". The parser in the first step throws away

the first occurrence of the term "software bugs" since it is followed by a hard stop, but retains the

second occurrence since it is followed by a soft stop. To prevent the parser from discarding good

noun phrases, such as the first occurrence of the term "software bugs", the second pass retrieves

all occurrences of the extracted phrases from the piece of text so that, for example, both

5   occurrences of the term "software bugs" are retrieved. Now, the first pass of the method will be

described in more detail.

    Figure 3B is a flowchart illustrating more details of the phrase extracting step 40 for

parsing a document in accordance with the invention. The method begins as a first word of the

document is loaded into the buffer from a document database or a memory of the server in step

10   42. Next, the parser determines if the word is a break character in step 44. The parser may also

delete certain characters or words at this stage of the parsing process. If the word is not a break

character, the method loops back to step 42 and the next word of the document is read into the

buffer. This process of reading a word into the buffer is repeated until a break character is

encountered so that the buffer contains a sequence of words (a phrase) which has a break

15   character before the sequence of words and a break character after the sequence of words. In this

manner, the document is parsed into phrases which are separated from one another by break

characters.

    If a break character is encountered, the parser may determine if the break character is an

explicit break character in step 46, delete the break character and extract the phrase contained in

20   the buffer if an explicit break character exists in step 48. The phrase extracted from the buffer

may be stored in a database for future use. Next, in step 50, the buffer may be flushed to empty

the words from the buffer and the buffer may begin loading new words into the buffer in steps 42

and 44 until another break character is identified. Returning to step 46, if the break character is

not an explicit break character, the parser determines if the break character is a soft stop word in

step 52. If the break character is a soft stop word, then the soft stop word is deleted and the

5   phrase in the buffer is stored in the database in step 54, the buffer is flushed in step 50 and the

buffer is refilled with new words from the document. If the break character is not a soft stop

word (i.e., the break character is a hard stop word), the hard stop word and the phrase in the

buffer are deleted in step 56, the buffer is flushed in step 50 and refilled with new words from the

document in steps 42 and 44. In this manner, phrases from the document are extracted in

10   accordance with the invention using the break characters and the type of break character to

separate the phrases from each other and determine which phrases are going to be saved in the

database. The parser in accordance with the invention does not attempt to analyze each word of

the document to identify key phrases as with conventional systems, but does extract phrases from

the document more quickly than conventional parsers and with as much accuracy as the

15   conventional parsers. Now, an example of the operation of the parser in accordance with the

invention will be described with reference to Figures 4 and 5A - 5L.

Figure 4 is an example of a document 60 to be parsed by the parsing system in

accordance with the invention while Figures 5A - 5L illustrate the operation of the buffer during

the parsing of the document 60 shown in Figure 4. In this example, the document is a short

20   electronic news story, but the parser may also extract phrases from any other piece of text. In

fact, the parser in accordance with the invention may be able to extract phrases from various

types of documents at speeds of up to 1 MByte of data per second. The particular story shown

describes a new "snake-like" robot developed by NEC. Figures 5A - 5L illustrate, in a table 68,

the operation of the buffer in accordance with the invention on the above story. In particular, a

first column 70 of the table contains the current word being read into the buffer, a second column

5   72 contains the determination of the type of word by the parser in accordance with the invention,

a third column 74 contains the contents of the buffer at the particular time, a fourth column 76

contains the word index (i.e., the phrases which are being extracted from the document) and a

fifth column 78 contains comments about the parsing process.

As shown in Figure 5A, the first word read into the buffer is a sequence of asterisks at the

10   beginning of the story which are classified by the parser as a break word (punctuation) and

deleted from the buffer. The next word is "Computer" which is entered into the buffer since it is

not a break word and the next word, which is "Select" is also entered into the buffer since it is

also not a break word. Thus, the buffer contains the phrase "Computer Select" as shown in a cell

80. The next word in the document is a comma which is classified as a break character by the

15   parser. Because the break character is punctuation (an explicit break), the words in the buffer are

saved in the database as shown in the Word Index column 76 and the buffer is flushed. Now,

new words are read into the buffer and parsed. The next word into the buffer is "October" which

is a hard stop word because it relates to a date and it is deleted. The next word received by the

buffer is "1995" which is a break character since it is a number and it is also deleted. The next

20   word received by the buffer is "COPYRIGHT" which is identified as a stop word because it is all

capital letters and it is deleted. The next word is "Newsbytes" which is not a break character and

is therefore stored into the buffer. The next word is "Inc." which is also stored in the buffer. The next word is a period which is a break character so that the buffer contents "Newsbytes Inc." are saved into a database as shown in the Word Index column, the break character is deleted and the buffer is flushed.

5   The next two word received by the buffer, which are "1995" and a sequence of asterisks, are both break words which are deleted. The next two words received by the buffer are "Newsbytes" and "Newsbytes" which are both stored in the buffer. The next word received is "August" which is a hard stop word so that the contents of the buffer and the hard stop word is deleted. The next three words received by the buffer are all break characters (i.e., numbers or

10   punctuation) which are deleted. The next word is a word containing a number in a cell 82 which is stored in the buffer, but then deleted when the next character is a break character because the buffer only contains a single word. As can be seen in Figures 5B - 5L, the parsing process continues for the entire document so that a list of key phrases, as shown in the Word Index column 76, are extracted from the document and saved in a database.

15   In summary, phrases which characterize the document or piece of text may be rapidly extracted from the document in accordance with the invention. The invention uses the break characters in the document or the piece of text to separate the phrases from each other and to extract the key phrases for a document. In the example above, the extracted phrases, such as "Newsbytes Inc.", "snake-like robot", "NEC Corporation", "robotically controlled electronic

20   snake", "disaster relief work" and "world's first active universal joint" permit a person reviewing only the key phrases to understand the context of the document without reviewing the entire

document. The parsing system in accordance with the invention performs the extraction of the key phrases more rapidly than any other conventional parsing systems which is important as the total amount of textual data and documents available for parsing increases at an exponential rate due, in part, to the explosion of the user of the Internet.

5   The parser in accordance with the invention may be used to parse documents in various different foreign languages with minor modifications to the rules database to reflect changes in the characters and changes in the syntax of the language. To better understand this, an example of a piece of Japanese text is described along with the resulting Japanese taxonomy. However, the invention may be used with a variety of different foreign languages with minor modifications
10   to the rules database.

          Figure 6 is an example of a piece of Japanese text 90 while Figure 7 illustrates a list 92 of phrases 94 that have been extracted from the piece of Japanese text using the two-pass parsing method in accordance with the invention.

          While the foregoing has been with reference to a particular embodiment of the invention,
15   it will be appreciated by those skilled in the art that changes in this embodiment may be made without departing from the principles and spirit of the invention, the scope of which is defined by the appended claims.

Claims:

1   1.   A system for parsing a piece of foreign language text into one or more phrases

2 which characterize a foreign language document, the system comprising:

3   a buffer for reading one or more words from the piece of text into the buffer until a break

4 character is identified;

5   a parser for identifying a phrase contained in the buffer, the phrase being a sequence of

6 two or more words in between break characters;

7   the parser further comprising means for determining the type of break character that

8 follows the identified phrase and means for saving a key phrase from the buffer based on the

9 determined type of break character ;

10   a database for storing the key foreign language phrases.

1   2.   The system of Claim 1, wherein the buffer further comprises means for flushing

2 the buffer when the key phrase is stored in the database or the phrase in the buffer is deleted.

1   3.   The system of Claim 1 further comprising a retriever for retrieving all occurrences

2 of the extracted phrases from the piece of text after the piece of text has been parsed.

1   4.   A method for parsing a piece of text into one or more phrases which characterize

2 the document, the method comprising:

3   reading one or more words from the piece of text into a buffer until a break character is

4 identified;

5   identifying a phrase contained in the buffer, the phrase being a sequence of two or more

6 words in between break characters;

7   determining the type of break character that follows the identified phrase; and

8   saving a key phrase from the buffer into a database based on the determined type of break

9   character.

1   5.   The method of Claim 4 further comprising flushing the buffer when the key

2   phrase is stored in the database or the phrase in the buffer is deleted.

1   6.   The method of Claim 4 further comprising retrieving all occurrences of the

2   extracted phrases from the piece of text after the piece of text has been parsed.

1   7.   A system for parsing a piece of text into one or more phrases which characterize a

2   document, the system comprising:

3   a buffer for reading one or more words from the piece of text into the buffer until a break

4   character is identified;

5   a parser for identifying a phrase contained in the buffer, the phrase being a sequence of

6   two or more words in between break characters;

7   the parser further comprising means for determining the type of break character that

8   follows the identified phrase and means for saving a key phrase from the buffer based on the

9   determined type of break character ;

10   a database for storing the key foreign language phrases; and

11   a retriever for retrieving all occurrences of the extracted phrases from the piece of text

12   after the piece of text has been parsed.

1   8.   The system of Claim 7, wherein the buffer further comprises means for flushing

2   the buffer when the key phrase is stored in the database or the phrase in the buffer is deleted.

1   9.   A method for parsing a piece of text into one or more phrases which characterize

2   the document, the method comprising:

3   reading one or more words from the piece of text into a buffer until a break character is

4   identified;

5   identifying a phrase contained in the buffer, the phrase being a sequence of two or more

6   words in between break characters;

7   determining the type of break character that follows the identified phrase;

8   saving a key phrase from the buffer into a database based on the determined type of break

9   character; and

10   retrieving all occurrences of the extracted phrases from the piece of text after the piece of

11   text has been parsed.

       10.   The method of Claim 9 further comprising flushing the buffer when the key

2   phrase is stored in the database or the phrase in the buffer is deleted.

1   11.   A system for parsing a piece of text into one or more phrases which characterize

2   the document, the system comprising:

3   a first pass comprising means for identifying a phrase contained in a buffer wherein the

4   phrase is a sequence of two or more words in between break characters, means for determining

5   the type of break character that follows the identified phrase and means for saving a key phrase

6   from the buffer based on the determined type of break character; and

7   a second pass comprising means for retrieving all occurrences of the extracted phrases

8   from the piece of text.

1   12.   A method for parsing a piece of text into one or more phrases which characterize

2   the document, the method comprising:

3   performing a first pass through the piece of text, the first pass comprising identifying a

4   phrase contained in a buffer wherein the phrase is a sequence of two or more words in between

5   break characters, determining the type of break character that follows the identified phrase and

6   saving a key phrase from the buffer based on the determined type of break character; and

7   performing a second pass through the piece of text comprising retrieving all occurrences

8   of the extracted phrases from the piece of text.

ABSTRACT OF THE DISCLOSURE

A parsing system and method are provided in which the break characters in the document are used to rapidly parse the document and extract one or more key phrases from the document which characterize the document. The break characters in the document may include explicit break characters, such as punctuation, soft stop words and hard stop words. The determination of which phrases in the document are extracted depends upon the type of break character appearing after the phrase in the document. The parser may also be used to parse a foreign language document into one or more phrases.
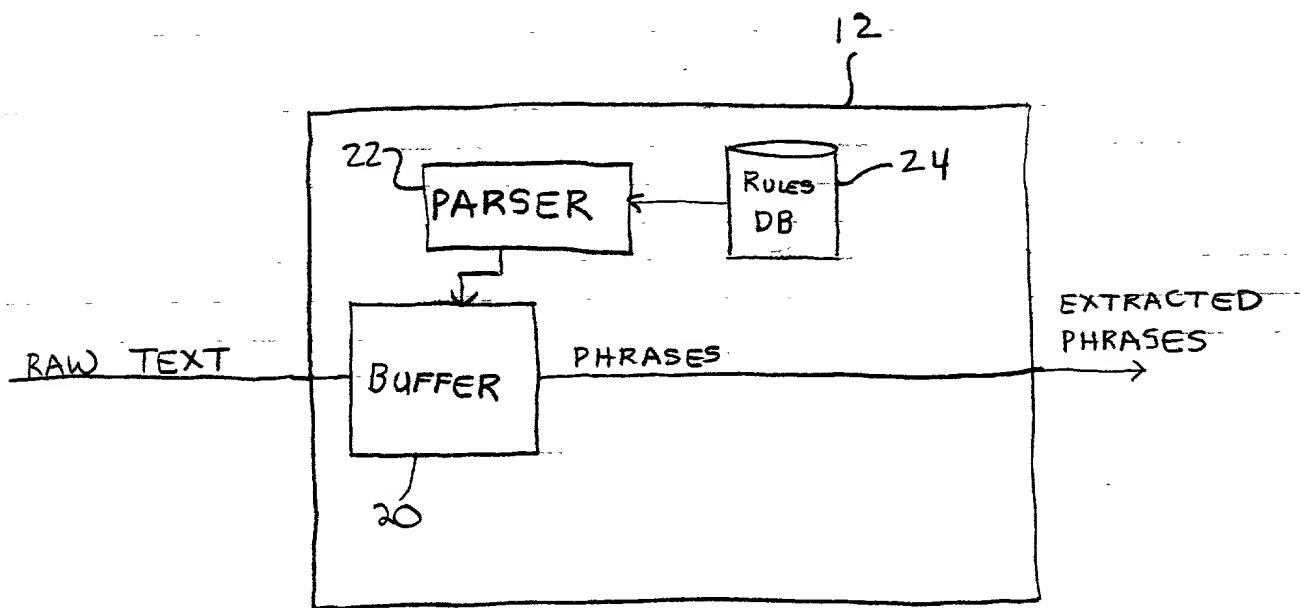
5

FIGURE 1

FIGURE 2

FIGURE 3A
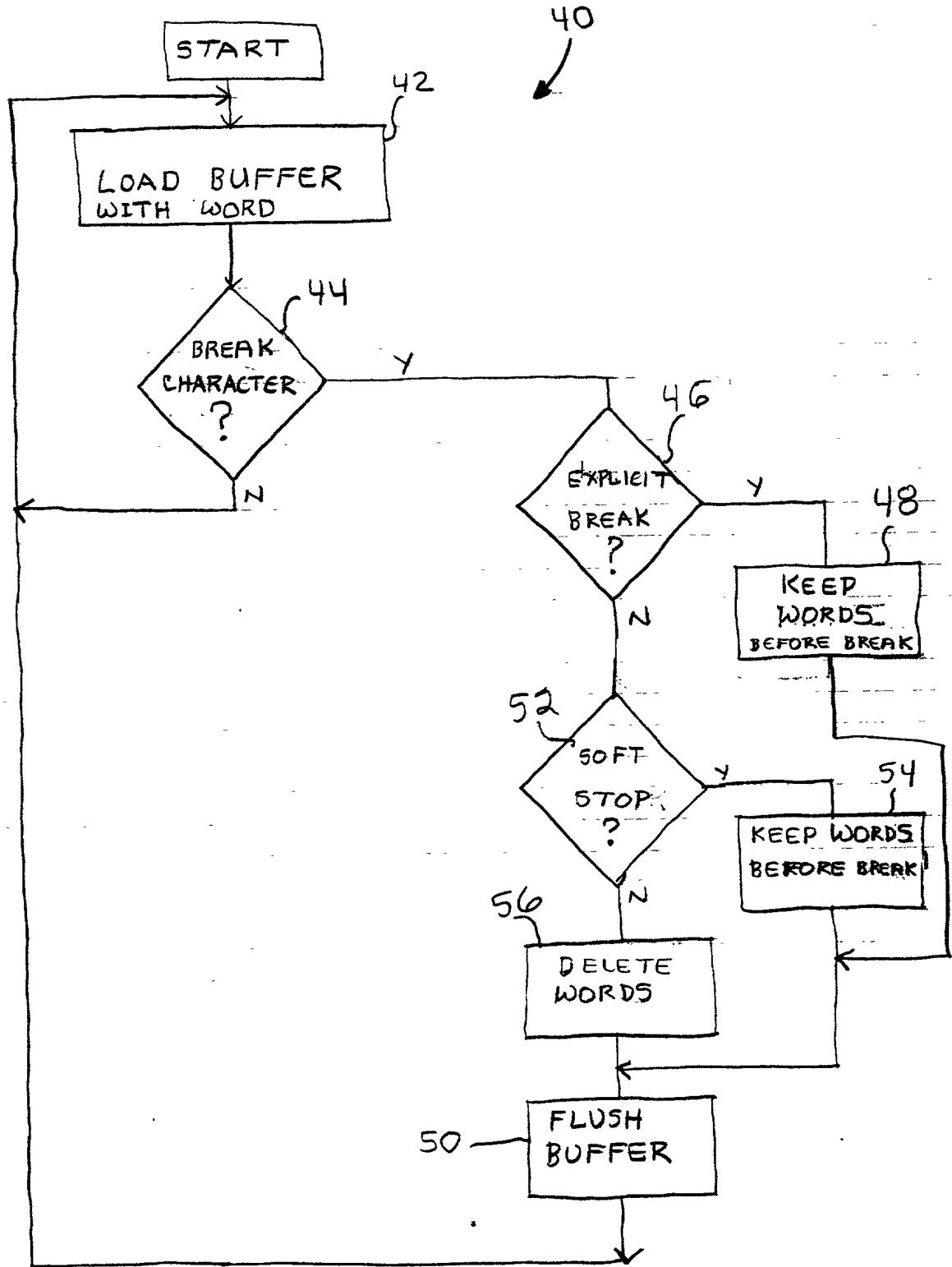
START

42

LOAD BUFFER
WITH WORD

44

BREAK
CHARACTER
?

Y

46

EXPLICIT
BREAK
?

Y

48

KEEP
WORDS
BEFORE BREAK

N

N

52

SOFT
STOP
?

Y

54

KEEP WORDS
BEFORE BREAK

N

56

DELETE
WORDS

50

FLUSH
BUFFER

40

FIGURE 3B

60

NEC Develops "Snake-Like" Robot.
Author
Mark, Jeremy
Full Text
TOKYO, JAPAN, 1995 AUG 31 (NB) -- NEC Corporation has developed a robotically controlled electronic snake that offers far more movement that previously designed robots. The company says that the much greater freedom of movement "makes it perfect" for everything from industrial to disaster relief work.

The secret of the new device lies in a revolutionary new type of universal joint. Previously joints in the body of robot have been restricted to movement in just one plane, either left and right or up and down. But the Tokyo-based company says it has succeeded in developing the world's first active universal joint.

Controlled by two motors, the joint allows full freedom of movement in all planes at each of the six joints along the robot's body, allowing the unit to crawl into places previously inaccessible.

At the heart of the robot is a computer processing unit that receives signals from the operator's handset and controls movement. The controller can instruct the computer to control all the joints in harmony or specify individual control of each joint if necessary.
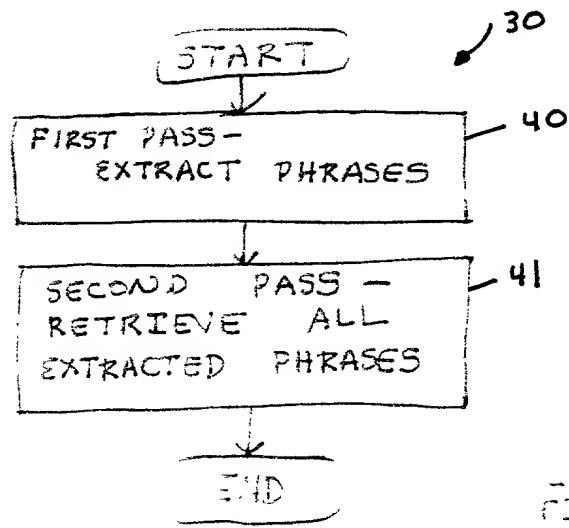
A video camera at the head of the robot sends signals back to the operator who can use them to steer the unit and also to examine places inaccessible to humans.

The entire device is 1.4 meters long and measures 42 millimeters in diameter. It weighs 4.6 kilograms.

The as-yet unnamed device is not yet commercially available, NEC's Mark Pearce told Newsbytes. "It will be a couple of years before everything in sorted out and it's ready to be sold. We have to increase the speed amongst other things," he said.

NEC says typical applications for such a robot could be investigation of complex pipework or as an aid to search teams in disaster hit areas where the device could crawl through the rubble of collapsed buildings.

(Martyn Williams/19950830/Press contact: Mark Pearce, NEC Corporation, tel +81-3-3798-6511, fax +81-3-3457-7249, Internet e-mail maku_10-22150@aladdin.nec.co.jp/NEC950831/PHOTO)

Record #
17 337 353

FIGURE 4

68

70 72 74 76 78

Word read	Word Type	Buffer	Word Index	Comments
******************************	Break			
Computer	Computer	Computer		
Select	Select	Computer Select		
,	Break	Flush	Computer Select	
October	Stop			
1995	Break			
COPYRIGHT	Stop			
Newsbytes	Newsbytes	Newsbytes		
Inc	Inc	Newsbytes Inc		
.	Break	Flush	Newsbytes Inc	Business markers
1995	Break			
******************************	Break			
Newsbytes	Newsbytes	Newsbytes		
Newsbytes	Newsbytes	Newsbytes Newsbytes		
August	Stop	Flush	Reject	
31	Break			
,	Break			
1995	Break			
pNEW08310001	pNEW08310001	pNEW08310001		Words containing numbers
-------------------------------	Break	Flush		
NEC	NEC	NEC		
Develops	active verb	Flush		
Snake-Like	Snake-Like	Snake-Like		
Robot	Robot	Snake-Like Robot		
.	Break	Flush	Snake-Like Robot	
Author	Author	Author		
CR	Break	Flush		
Mark	Mark	Mark		
,	Break	Flush		
Jeremy	Jeremy	Jeremy		
CR	Break	Flush		
Full	Full	Full		
Text	Text	Full Text		
CR	Break	Flush	Full Text	
TOKYO	TOKYO	TOKYO		

80

82

FIGURE 5A

Word read	Word Type	Buffer	Word Index	Comments
,	Break	Flush		
JAPAN	JAPAN	JAPAN		
,	Break	Flush		
1995	Break			
AUG	Stop			
31	Break			
(	Break			
NB	Stop			
)	Break			
--	Break			
NEC	NEC	NEC		
Corporation	Corporation	NEC Corporation		
has	Stop	Flush	NEC Corporation	
developed	developed	developed		
a	Stop	Flush		
robotically	robotically	robotically		
controlled	controlled	robotically controlled		
electronic	electronic	robotically controlled electronic		
snake	snake	robotically controlled electronic snake	robotically controlled electronic snake	
that	Stop	Flush		
offers	active verb			
far	Stop			
more	Stop			
movement	movement	movement		
that	Stop	Flush		
previously	previously	previously		
designed	designed	previously designed		
robots	robots	previously designed robots		
.	Break	Flush	previously designed robots	
The	Stop			
company	company	company		
says	Stop	Flush		
that	Stop			
the	Stop			

FIGURE 5B

Word read	Word Type	Buffer	Word Index	Comments
much	Stop			
greater	Stop			
freedom	freedom	freedom		
of	Connector	freedom of		
movement	movement	freedom of movement		
makes	Stop	Flush	freedom of movement	
it	Stop			
perfect	perfect	perfect		
for	Stop	Flush		
everything	Stop			
from	Stop			
industrial	industrial	industrial		
to	Stop	Flush		
disaster	disaster	disaster		
relief	relief	disaster relief		
work	work	disaster relief work		
.	Break	Flush	disaster relief work	
The	Stop			
secret	secret	secret		
of	Connector	secret of		
the	THE	secret of the		
new	new	secret of the new		
device	device	secret of the new device		
lies	active verb	Flush	secret of the new device	
in	Stop			
a	Stop			
revolutionary	revolutionary	revolutionary		
new	new	revolutionary new type		
type	type	revolutionary new type		
of	Connector	revolutionary new type of		
universal	universal	revolutionary new type of universal		
joint	joint	revolutionary new type of universal joint		

FIGURE 5C

Word read	Word Type	Buffer	Word Index	Comments
.	Break	Flush	revolutionary new type of universal joint	
Previously	Previously	Previously		
joints	active verb	Flush		
in	Stop			
the	Stop			
body	body	body		
of	Connector	of		
robot	robot	body of robot		
have	Stop	Flush	body of robot	
been	Stop			
restricted	restricted	restricted		
to	Stop	Flush		
movement	movement	movement		
in	Stop	Flush		
just	Stop			
one	Stop			
plane	plane	plane		
,	Break	Flush		
either	Stop			
left	left	left		
and	Connector	left and		
right	right	left and right		
or	Stop	Flush	left and right	
up	Stop			
and	Connector			
down	Stop			
.	Break			
But	Stop			
the	Stop			
Tokyo-based	Tokyo-based	Tokyo-based		
company	company	Flush company		
says	Stop	Flush		
it	Stop			
has	Stop			
succeeded	succeeded	succeeded		
in	Stop	Flush		
developing	developing	developing		
the	Stop	Flush		
world's	world's	world's		
first	first	world's first		

FIGURE 5D

Word read	Word Type	Buffer	Word Index	Comments
active	active	world's first active		
universal	universal	world's first active universal		
joint	joint	world's first active universal joint		
.	Break	Flush	world's first active universal joint	
Controlled	Controlled	Controlled		
by	Stop	Flush		
two	Stop			
motors	motors	motors		
,	Break	Flush		
the	Stop			
joint	joint	joint		
allows	active verb	Flush		
full	full	full		
freedom	freedom	full freedom		
of	Connector	full freedom of		
movement	movement	full freedom of movement		
in	Stop	Flush	full freedom of movement	
all	Stop			
planes	active verb			
at	Stop			
each	Stop			
of	Connector			
the	THE			
six	Stop			
joints	active verb			
along	Stop			
the	Stop			
robot's	robot's	robot's		
body	body	robot's body		
,	Break	Flush	robot's body	
allowing	allowing	allowing		
the	Stop	Flush		
unit	unit	unit		
to	Stop	Flush		
crawl	crawl	crawl		

FIGURE 5E

Word read	Word Type	Buffer	Word Index	Comments
into	Stop	Flush		
places	active verb			
previously	previously	previously		'
inaccessible	inaccessible	previously inaccessible		
.	Break	Flush	previously inaccessible	
At	Stop			
the	Stop			
heart	heart	heart		
of	Connector	heart of		
the	THE	heart of the		
robot	robot	heart of the robot		
is	Stop	Flush	heart of the robot	
a	Stop			
computer	computer	computer		
processing	processing	computer processing		
unit	unit	computer processing unit		
that	Stop	Flush	computer processing unit	
receives	active verb			
signals	signals	signals		
from	Stop	Flush		
the	Stop			
operator's	operator's	operator's		
handset	handset	operator's handset		
and	Connector	operator's handset and		
controls	active verb	Flush	operator's handset	
movement	movement	movement		
.	Break	Flush		
The	Stop			
controller	controller	controller		
can	Stop	Flush		
instruct	instruct	instruct		
the	Stop	Flush		
computer	computer	computer		
to	Stop	Flush		
control	control	control		

FIGURE 5F

Word read	Word Type	Buffer	Word Index	Comments
all	Stop	Flush		
the	Stop			
joints	active verb			
in	Stop			
harmony	harmony	harmony		
or	Stop	Flush		
specify	specify	specify		
individual	individual	individual		
control	control	individual control		
of	Connector	individual control of		
each	Stop	Flush	individual control	
joint	joint	joint		
if	Stop	Flush		
necessary	necessary	necessary		
.	Break	Flush		
A	Stop			
video	video	video		
camera	camera	video camera		
at	Stop	Flush	video camera	
the	Stop			
head	head	head		
of	Connector	head of		
the	THE	head of the		
robot	robot	head of the robot		
sends	active verb	Flush	head of the robot	
signals	signals	signals		
back	Stop	Flush		
to	Stop			
the	Stop			
operator	operator	operator		
who	Stop	Flush		
can	Stop			
use	Stop			
them	Stop			
to	Stop			
steer	steer	steer		
the	Stop	Flush		
unit	unit	unit		
and	Connector	unit and		
also	Stop	Flush		
to	Stop			

FIGURE 5G

Word read	Word Type	Buffer	Word Index	Comments
examine	examine	examine		
places	active verb	Flush		
inaccessible	inaccessible	inaccessible		
to	Stop	Flush		
humans	humans	humans		
.	Break	Flush		
The	Stop			
entire	entire	entire		
device	device	entire device		
is	Stop	Flush	entire device	
1	Break			
.	Break			
4	Stop			
meters	active verb			
long	long	long		
and	Connector	long and		
measures	active verb	Flush		
42	Break			
millimeters	millimeters	millimeters		
in	Stop	Flush		
diameter	diameter	diameter		
.	Break	Flush		
It	Stop			
weighs	weighs	weighs		
4	Break	Flush		
.	Break			
6	Break			
kilograms	kilograms	kilograms		
.	Break	Flush		
The	Stop			
as-yet	as-yet	as-yet		
unnamed	unnamed	as-yet unnamed		
device	device	as-yet unnamed device		
is	Stop	Flush	as-yet unnamed device	
not	Stop			
yet	Stop			
commercially	commercially	commercially		
available	available	commercially available		
,	Break	Flush	commercially available	

FIGURE 5H

Word read	Word Type	Buffer	Word Index	Comments
NEC's	NEC's	NEC's		
Mark	Mark	NEC's Mark		
Pearce	Pearce	NEC's Mark Pearce	NEC's Mark Pearce	
told	active verb (past tense)	Flush told		
Newsbytes	Newsbytes	Flush Newsbytes		
.	Break	Flush		
It	Stop			
will	Stop			
be	Stop			
a	Stop			
couple	couple	couple		
of	Connector	couple of		
years	years	couple of years		
before	Stop	Flush	couple of years	
everything	Stop			
in	Stop			
sorted	sorted	sorted		
out	Stop	Flush		
and	Connector			
it's	Stop			
ready	ready	ready		
to	Stop	Flush		
be	Stop			
sold	sold	sold		
.	Break	Flush		
We	Stop			
have	Stop			
to	Stop			
increase	increase	increase		
the	Stop	Flush		
speed	speed	speed		
amongst	Stop	Flush		
other	Stop			
things	things	things		
,	Break	Flush		
he	Stop			
said	Stop			
.	Break			
NEC	NEC	NEC		
says	Stop	Flush		
typical	typical	typical		

FIGURE 5I

Word read	Word Type	Buffer	Word Index	Comments
applications	applications	typical applications		
for	Stop	Flush	typical applications	
such	Stop			
a	Stop			
robot	robot	robot		
could	Stop	Flush		
be	Stop			
investigation	investigation	investigation		
of	Connector	investigation of		
complex	complex	investigation of complex		
pipework	pipework	investigation of complex pipework		
or	Stop	Flush	investigatio n of complex pipework	
as	Stop			
an	Stop			
aid	aid	aid		
to	Stop	Flush		
search	search	search		
teams	active verb	Flush		
in	Stop			
disaster	disaster	disaster		
hit	hit	disaster hit		
areas	areas	disaster hit areas		
where	Stop	Flush	disaster hit areas	
the	Stop			
device	device	device		
could	Stop	Flush		
crawl	crawl	crawl		
through	Stop	Flush		
the	Stop			
rubble	rubble	rubble		
of	Connector	rubble of		
collapsed	collapsed	rubble of collapsed		

FIGURE 5J

Word read	Word Type	Buffer	Word Index	Comments
buildings	buildings	rubble of collapsed buildings		
.	Break	Flush	rubble of collapsed buildings	
(	Break			
Martyn	Martyn	Martyn		
Williams/199508 30/Press	Williams/19950 830/Press	Martyn Williams/1995 0830/Press	Martyn Williams/19 950830/Pre ss	Words containing numbers
contact	contact	Flush contact		
:	Break	Flush		
Mark	Mark	Mark		
Pearce	Pearce	Mark Pearce		
,	Break	Flush	Mark Pearce	
NEC	NEC	NEC		
Corporation	Corporation	NEC Corporation		
,	Break	Flush	NEC Corporation	
tel	Stop			
+81-3-3798-6511	Break			
,	Break			
fax	Stop			
+81-3-3457-7249	Break			
,	Break			
Internet	Internet	Internet		
e-mail	Stop	Flush		
maku_10- 22150@aladdin	maku_10- 22150@aladdin	maku_10- 22150@aladdi n		Words containing numbers Words containing Internet codes
.	Break	Flush		
nec	nec	nec		
.	Break	Flush		
co	co	co		
.	Break	Flush		
jp/NEC950831/P HOTO	jp/NEC950831/ PHOTO	jp/NEC95083 1/PHOTO		Words containing numbers
)	Break	Flush		
--------------------- ---------	Break			

FIGURE 5K

Word read	Word Type	Buffer	Word Index	Comments
Record	Record	Record		
#	Break	Flush		
17	Break			
337	Break			
353	Break			

FIGURE 5L

第1部 環境行政の課題と展開

第1節 環境行政の現状と課題

　大量生産と大量消費に基づく今日の社会経済システムは、人々の快適な生活を支える一方で、エネルギーの消費とあいまって、今や国際課題とも言うべき、地球温暖化をはじめとする様々な環境問題を引き起こす結果をもたらした。

　また、本県においても、都市中小河川・湖沼の水質汚濁や自動車排出ガスによる大気汚染などの都市・生活型公害が顕在化しているほか、近年、ごみの増加に伴い発生するダイオキシン類や新たな環境ホルモンの問題、廃棄物処理とリサイクル対策など、克服しなければならない課題が発生してきている。

　一方、人々の価値観が多様化し、生活の質をより重視する時代へと変化してきていることから、豊かな生活環境や自然とのふれあいの場など、心安らぐ、より質の高い地域環境の保全・創造が求められている。

　こうした広範な環境問題に適切に対応するためには、総合的な視点に立って、各種施策を計画的に、また、着実に進めることが必要となっている。

1　有害化学物質による環境汚染の防止

　科学技術の進展に伴い、生産・使用される化学物質は多種にわたり、また、その排出形態も多様化している。さらに、ダイオキシン類など非意図的に生成するものもあり、有害化学物質による環境汚染への対応が大きな課題となっている。

　私たちを取り巻く環境は、大気・水・土壌・生物等を構成要素として成り立っているが、生産、使用、廃棄等の各段階で排出される化学物質は、これらの要素の間を循環することになる。こうした循環の過程を通じて、人間はもとよりあらゆる生物が様々の化学物質にさらされているが、一部の化学物質については、極めて低濃度でも生物に影響を及ぼすおそれがあるにもかかわらず、その有害性の程度はもとより、大気・水・土壌等環境中での挙動についての科学的知見は必ずしも十分とはいえない状況にある。

　特に、環境ホルモンと呼ばれる化学物質については、国において調査研究が緒に就いた段階であり、環境ホルモンとして内分泌攪乱作用が疑われている約70の物質についてすら、人類をはじめ野生生物等に対する作用のメカニズムは未解明である。また、多くの化学物質について、その環境への排出量や廃棄物等としての移動量などの実態はいまだ把握されていないのが現状である。

Figure 6

福井県環境白書
環境行政の課題と展開
環境行政の現状と課題
大量生産と大量消費
今日の社会経済システム
快適な生活
エネルギーの消費
国際課題
地球温暖化
環境問題
都市中小河川
湖沼の水質汚濁
自動車排出ガス
大気汚染
生活型公害
ダイオキシン類
環境ホルモンの問題
廃棄物処理とリサイクル対策
生活の質
生活環境
地域環境の保全
広範な環境問題
総合的な視点
各種施策
有害化学物質
環境汚染の防止
科学技術の進展
化学物質
排出形態

92

94

Figure 7

DECLARATION AND POWER OF ATTORNEY

DECLARATION:

As a below named inventor, I hereby declare that:

My residence, post office address and citizenship are as stated below next to my name.

I believe, I am the original, first and sole inventor (if only one name is listed below) or an original, first and joint inventor (if plural names are listed below) of the subject matter which is claimed and for which a patent is sought on the invention entitled:

SYSTEM AND METHOD FOR PARSING A DOCUMENT

the specification of which (check only one item below):

    <u>X</u>   is attached hereto.

    ___   was filed as United States Application
    Serial No._____ on _____
    and was amended on _____ (if applicable).

___ was filed as PCT international application
    Number _____ on _____
    and was amended under PCT Article 19
    on _____ (if applicable).

I hereby state that I have reviewed and understand the contents of the above-identified specification, including the claims, as amended by any amendment referred to above.

I acknowledge the duty to disclose information which is material to the examination of this application in accordance with Title 37, Code of Federal Regulations, §1.56(a).

I hereby claim foreign priority benefits under Title 35, United States Code, §119 of any foreign application(s) for patent or inventor's certificate or of any PCT international application(s) designating at least one country other than the United States of America listed below and have also identified below any foreign application(s) for patent or inventor's certificate or any PCT international application(s) designating at least one country other than the United States of America filed by me on the same subject matter having a filing date before that of the application(s) on which priority is claimed:

PRIOR FOREIGN/PCT APPLICATION(S) AND ANY PRIORITY CLAIMS UNDER 35 U.S.C. 119:			
Country (If PCT, indicate PCT)	Application Number	Date Filed	Priority Claimed (Yes/No)

I hereby claim the benefit under Title 35, United States Code, §120 of any United States application(s) or PCT international application(s) designating the United States of America that is/are listed below and, insofar as the subject matter of each of the claims of this application is not disclosed in that/those prior application(s) in the manner provided by the first paragraph of Title 35, United States Code, §112, I acknowledge the duty to disclose material information as defined in Title 37, Code of Federal Regulations, §1.56(a) which occurred between the filing date of the prior application(s) and the national or PCT international filing date of this application:

PRIOR U.S. APPLICATIONS OR PCT INTERNATIONAL APPLICATIONS DESIGNATING THE U.S. FOR BENEFIT UNDER 35 U.S.C. 120:				
U.S. APPLICATIONS		STATUS (check one)		
U.S. APPLICATION NUMBER	U.S. FILING DATE	PATENTED	PENDING	ABANDONED
09/288,994	April 9, 1999		X	
PCT APPLICATIONS DESIGNATING THE U.S.				
PCT APPLICATION NO.	PCT FILING DATE	U.S. SERIAL NUMBERS ASSIGNED (if any)		

POWER OF ATTORNEY:

As a named inventor, I hereby appoint the following attorney(s) and/or agent(s) with full power of substitution to act exclusively to prosecute this application and transact all business in the Patent and Trademark Office connected therewith:

Barry N. Young (Reg. No. 27,744); Timothy W. Lohse (Reg. No. 35,255); Stephen E. Reiter (Reg. No. 31,192); Steven R. Sprinkle (Reg. No. 40,825); William N. Hulsey III (Reg. No. 33,402); Terrance A. Meador (Reg. No. 30,298); Ramsey R. Stewart (Reg. No. 38,322); June M. Learn (Reg. No. 31,238); John Oskorep (Reg. No. 41,234); Timothy N. Ellis (Reg. No. 41,734); William G. Goldman (Reg. No. 42,590); Sheila Kirschenbaum (Reg. No. 44,835); Travis L. Dodd (Reg. No. 42,491); Charles D. Gavrilovich, Jr. (Reg. No. 41,031); Gerald W. Maliszewski (Reg. No. 38054); Hayward A. Verdun (Reg. No. 43,223); Armando Pastrana, Jr. (Reg. No. 44,997); Richard M. Goldman (Reg. No. 25,585)

All correspondence should be addressed to:

Timothy W. Lohse
GRAY CARY WARE & FREIDENRICH
Patent Department - Hillview
3340 Hillview Avenue
Palo Alto, CA 94304

All telephone calls should be directed to Timothy W. Lohse, telephone number (650) 320-7400.

Inventor's Full Name:	Claude VOGEL
Inventor's Signature:	
Date:	
Residence: (City, State and/or country)	94140 Alfortville, France
Citizenship:	France
Post Office Address:	21, Rue Raymond Jaclard